IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 13, 2018, accepted October 17, 2018, date of publication November 9, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2879156

Structural Regression Model Based Inverse Sparse Representation for Tracking Objects

XIAOWEI AN[1], NONGLIANG SUN[2], AND MAOYONG CAO[1]
[1]College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China
[2]College of Electronics, Communication and Physics, Shandong University of Science and Technology, Qingdao 266590, China

Corresponding authors: Nongliang Sun (nl-jackson@vip.163.com) and Maoyong Cao (my-cao@263.net)

This work was supported in part by Leading Talents of Shandong University of Science and Technology, in part by the 863 Sub-project Verification Platform for Dynamic Evolution Technology of Mine Disaster under Grant 2015AA016404-4, in part by the Shandong Province Higher Educational Science and Technology Program under Grant J17KA075, and in part by the National Nature Science Foundation of China under Grant 61801270.

ABSTRACT In order to reduce the calculation cost and improve the robustness of appearance model, this paper presents an optimal object tracking method that consists of improved inverse sparse representation and global spatial envelope. First, partial least squares regression-based structural model is adopted, which easily facilitates target template sparsely represented by candidate dictionary. Furthermore, candidates with nonzero coefficients are easily selected as possible tracking results. Meanwhile, partial occlusion and slight appearance changes are effectively alleviated during the tracking process. Second, spatial envelope in the frequency domain is utilized to select the best candidate from the inverse sparse representation process. Multiple scales and orientations-based Gabor filters are established to obtain the Gist information, which keeps the potential structural attributes of local appearance models to tolerate appearance variation. In addition, the Bayesian inference framework is used to exploit candidate samples, and a simple model update scheme is employed to alleviate drifting caused by temporal varying multi-factors. The qualitative experimental results show that the proposed tracking algorithm provides a better performance in some dynamic scenes.

INDEX TERMS Optimal appearance model, partial least squares regression, inverse sparse representation, Gist.

I. INTRODUCTION

Object tracking is a major research direction in computer vision. Several algorithms have been employed for numerous vision applications in the past years [1]–[3]. The main challenges of tracking process are partial occlusion, fore/background clutter, illumination changes, pose and scale variation. Therefore, it is a crucial task to describe the model to avoid similar influences [2].

Generally, a tracking method consists of three important parts: motion model, observation model and localization strategy within the whole tracking process. The motion model provides the most similar candidate sets for consecutive matching. The observation model measures the likelihood of the possible candidates, and the localization strategy updates the observation model adaptively according to the variations of target appearance.

Numerous model schemes have given different explanations in this scenario. In [4], the incremental visual tracker (IVT) developed a subspace model that could deal with the appearance changes based on principle component analysis. Method presented in [5] treated visual tracking as a multi-task sparse subspace learning problem. Yang *et al.* [6] proposed a tracking method based on super-pixels (SPT) which kept the potential structural information in the local model. Liu *et al.* [7] proposed the statistical representation based on meanshift model which constructed discriminative patch-wise sparsity histogram. Zhong [8] proposed a fusion model which consisted of discriminative and generative features that treated the observation model as sparsity representation. ASLA [9] combined the spatial information with a novel alignment-pooling strategy to model the target appearance. Sevilla-Lara [10] took distribution fields (DFs) to model the target appearance which resulted from the gradient descent method. In order to improve the target appearance robustness, [11] and [12] used the trivial patches to make the sparse dictionary. Wang *et al.* [13] and Bai *et al.* [14]

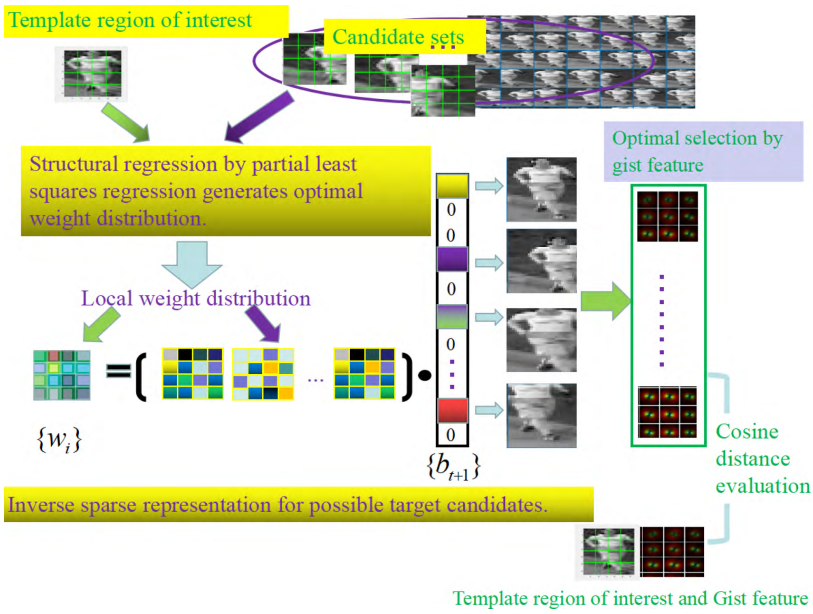


FIGURE 1. Illustration of the proposed algorithm.

exploited patch-based regression model to represent the local appearance. Through L_1 minimization, the responses to the base-vectors were checked to identify the object image patches. But the aforementioned applications always demanded much computation cost in real-time process. Slow tracking speed is still inferred by the multiple sparse decompositions. To reduce the computation cost, [15] proposed the inverse sparse representation where template was coded by the candidate sets, and the coefficients gave the measurement of similarity between the candidates and templates. However the update strategy for local appearance model still could not avoid complex boosting procedures.

Recently, the convolution neural network (CNN) based methods and its variants have attracted more attention in visual process. In [16], a compact structure was constructed by multi-layers in order to facilitate the extraction of discriminative information in the tracking process. Zhang *et al.* [17] presented effective feature distribution to avoid complex calculation in the data pre-training. However, it is still low calculation speed in the tracking. Meanwhile, large sets of pre-training processes were needed that labeling and training ground-truths data always spend complex-high cost in the visual data collection. Wang and Ge [18] combined inverse sparse representation and the double-structural network that easily labeled all candidate samples simultaneously. Inspired by Wang *et al.* [13], Bai *et al.* [14], Wang *et al.* [15], and Wang *et al.* [18], to improve the robustness of observation model and to avoid the complex calculation cost of the model, this paper proposes a novel tracking method based on the inverse sparse model with optimal structural regression.

In this paper, the least absolute shrinkage and selection operator(LASSO) based sparse decomposition is avoided frequently being used in each loop because it wastes too much computation cost. After localization of possible candidates, partial least squares regression exploits the potential information of local patch weight distribution which facilitate to update dictionary. Moreover, to alleviate the drifts, Gist descriptor from spatial envelope is employed to select the optimal candidate choice in the tracking model. Fig. 1 illustrates the proposed algorithm. The proposed method offers several advantages in the tracking process as follows:

1). It offers a novel representative structural regression model based inverse sparse representation(SRMISR).

2). Structural regression model by partial least squares for inverse sparse representation largely saves computation cost and exploits the potential structural information within local variant appearance model so that it is able to detect the target accurately and can run in real-time.

3). Spatial envelope is employed into this work that takes a novel evaluation scheme for selecting the best candidate. Multiple orientations and scales based gabor filters are adopted to extract the target representation, which can preserve the structure features of candidates more effective.

The rest of the paper is organized as follows: Section II presents the preliminary related knowledge. Section III presents the details of the proposed algorithm which combines inverse sparse representation and utilization of new local patch-weights distribution that results from partial least squares regression. Meanwhile, we present a novel way that facilitates candidates selection. Section IV presents comparative experiments that prove the effectiveness and efficiency of our algorithm. Section V concludes with a discussion of the results and recommendations for the future work.

II. BACKGROUND INFORMATION

This section first briefly introduces the bayesian inference tracking framework. Then we present some important notations used in this paper.

A. BAYESIAN INFERENCE TRACKING FRAMEWORK

Given consecutive video frames, let X_t denotes the state vector which describes the target motion variables. In this paper, affine transformation is employed to describe six state variables $X_t = \{x_t, y_t, \theta_t, s_t, \delta_t, \phi_t\}$: denoting x, y *transition, rotation angle, scale, aspect ratio, skew direction* at time t. [19]

let $\{Z_t\}$ denotes the corresponding observation vectors. The bayesian inference can be described as following (1):

$$p(X_t|Z_{1:t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|Z_{1:t-1})dX_{t-1}$$
$$p(X_t|Z_{1:t}) = \frac{p(Z_t|X_t)p(X_t|Z_{1:t-1})}{p(Z_t|Z_{1:t-1})} \quad (1)$$

Where $X_{1:t} = \{X_i\}_{i=1:t}$ represent target motion state vectors up to t-th time and $Z_{1:t} = \{Z_i\}_{i=1:t}$ stand for the corresponding observations. $p(X_t|X_{t-1})$ is the state transition model between the reference frame and candidate frame.

Generally, particle filter treats the posterior $p(X_t|Z_{1:t})$ as N weighted sampling particles $\{X_t^i, \boldsymbol{\beta}_t^i\}_{i=1,\ldots,N}$. According to the different global weights $\{\boldsymbol{\beta}_t^i\}_{i=1,\ldots,N}$ in the distribution, the new state $\widehat{X}_t = \sum_{i=1}^{N} \boldsymbol{\beta}_t^i X_t^i$ can be predicted.

B. INVERSE SPARSE REPRESENTATION FORMULATION

M candidate states $\{X_{t+1}^M\}$ given by particle filtering at the $(t+1)$-th frame can be obtained in the previous target region of interest(ROI) neighborhood of t-th frame. Observation matrices from the candidate sets $\{X_{t+1}^M\}$ can be coded for dictionary $\{D_{t+1}\} = [X_{t+1}^1, X_{t+1}^2, \ldots, X_{t+1}^M] \in R^{d\times M}$.

Afterwards, sparse decomposition of template $t_t \in R^{d\times 1}$ is presented by non-negative combination of sparse coefficients $\{b_{t+1}^*\} = [b_{t+1}^1, b_{t+1}^2, \ldots, b_{t+1}^M] \in R^{1\times M}$ while template reconstruction error achieves the minimum constraint with penalty term λ under optimal local patch weight distribution w^{t+1} as shown in (2):

$$arg\ \min ||w^{t+1} \odot (t_t - D_{t+1}b_{t+1}^*)|| + \lambda||b_{t+1}^*||_1$$
$$s.t.\ b_{t+1}^* \geqq 0 \quad (2)$$

Fig. 2 shows that ROI in the template frame can be represented by the global samples from the candidate frame with sparse combination coefficients$\{b_{t+1}^*\}$.

C. PARTIAL LEAST SQUARES REGRESSION MECHANISM

Partial least squares regression (PLSR) is the optimal statistical learning tool for describing the correlation among observation sets through the estimation of a low dimensional latent space which maximizes the separation between samples with different characteristics [20]. The PLSR builds new predictive variables which are called latent variables that make a link between the $n \times p$ matrix G of features and the $n \times q$ vector H of the response class labels.

The general regression of multivariate PLSR [21] is as shown in (3)

$$\min ||H_{nq} - G_{np}W_{pq}|| \quad (3)$$

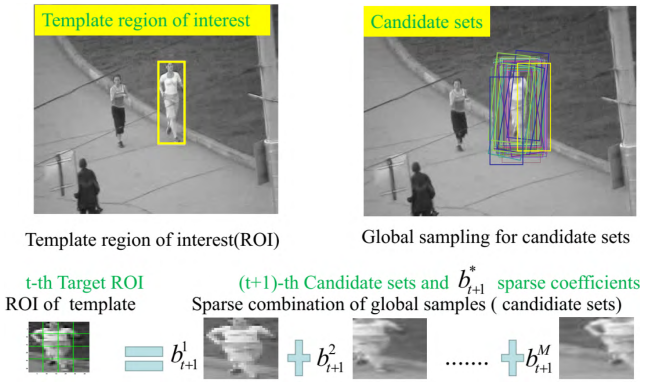


FIGURE 2. Inverse sparse representation process.

Here W_{pq} is the regression coefficient matrix with $p \times q$ size. Actually G and H are decomposed as shown in (4):

$$G_{np} = T_{nr}P_{pr}{}^T + E_{np}$$
$$H_{nq} = U_{nr}Q_{qr}{}^T + F_{nq} \quad (4)$$

Here T_{nr} and U_{nr} are $n \times r$ size low-dimensional latent representation of G and H. P_{pr} and Q_{qr} are the matrices of loadings. E_{np} and F_{nq} are residuals. According to the nonlinear iterative partial least squares (NIPALS) algorithm [20], PLS is able to calculate the weight vectors w_i as in (5):

$$\max(COV(t_i, u_i)) = \max(COV(H, Gw_i)) \quad (5)$$

Here t_i and u_i are the corresponding i-th column vectors in T_{nr} and U_{nr}, respectively. $\{w_i\}$ is the i-th column vectors in the set of weight vector W_{pq}. Afterwards, iterative calculation about the obtained column vector t_i and u_i is shown in (6), also p_i and q_i are the corresponding i-th column vectors in P and Q, respectively. Then the G and H are able to be iteratively denoted as (7)

$$p_i = (G^T t_i)/{t_i}^T t_i$$
$$q_i = (H^T u_i)/{u_i}^T u_i \quad (6)$$
$$G \leftarrow G - t_i G^T$$
$$H \leftarrow H - u_i H^T \quad (7)$$

If the residual $\min ||H_{nq} - G_{np}W_{pq}||$ is smaller than the setting threshold in the (7) iteration, the latent factors vectors w_{pq} is obtained completely.

D. SPATIAL ENVELOPE BASED GIST DESCRIPTOR

Spatial envelope with different particular frequencies and orientations describe the spatial frequency structure in images effectively that preserve latent relations, thus patterns of orientation-dependent frequency are extracted fluently [22]. According to the frequency information captured by the convolutional process, the representation of target candidate is enhanced more. Gabor convolutional kernel has been identified as the most similar profile of cortical simple cell receptive fields. It shows that effective characteristics of selectivity and locality which are optimal in the spatial envelope domain.

FIGURE 3. Selection of positive and negative labels; solid rectangle labels the filtering optimal state that is also the positive case for PLS. dashed rectangles are negative cases for partial least squares regression.

Impulse response G for gabor filtering processing is denoted in [23].

$$G(r, c, \sigma, \theta) = \frac{1}{2\pi\sigma^2}\exp(-\frac{r^2+c^2}{2\sigma^2}) \times \exp(j2\pi(r\cos\theta + c\sin\theta)) \quad j = \sqrt{-1} \tag{8}$$

where (r, c) represents a pixel position in the image, σ and θ represent the bandwidth and the orientation of gabor filter, respectively. The Gist descriptor is denoted by filtering the image based on banks of gabor filters. In other words, Gist descriptor is a combination of gabor descriptors with multiple directions and scales.

III. PROPOSED ALGORITHM

The main contribution of this work is presented in this section. A generic approach for incorporating partial least squares regression and inverse sparse representation is presented. A novel candidate selection scheme for the final target localization guided by the Gist descriptor governing optimal global physical attributes of the object is proposed and described within details.

A. STRUCTURAL REGRESSION MODEL BASED INVERSE SPARSE REPRESENTATION(SRMISR)

After ROIs sampled from dynamic scene, each normalized sample from template and candidate sets will be segmented into non-overlapping K parts that facilitate sparse decomposition as shown in Fig. 4. Template region of interest $t_t \in R^{d\times 1}$ is manually labeled groundtruth in the t-th frame. The consecutive candidate image is factorized into column vectors then normalized by initial weights following the uniform distribution. So each local patch in the column vectors can be adaptively adjusted in subsequent tracking process to maintain enough structural information which exist inside local appearance model.

For the sparse decomposition, based on the dictionary subsets $\{D_{t+1}\}$, the selected non-zero coefficients b^*_{t+1} compactly represents the same attributes as input signal $t_t \in R^{d\times 1}$. Particle filtering optimal state X^{optm}_{t+1} in the $(t+1)$-th frame under the inverse sparse representation can be denoted in (9) and shown as Fig. 2

$$X^{optm}_{t+1} = \sum_{i=1}^{M} b^i_{t+1} X^i_{t+1} \quad s.t.\ b^i_{t+1} \geq 0 \tag{9}$$

Once X^{optm}_{t+1} has been obtained in the candidate frame, some positive ROIs and negative ROIs can be easily established according to the filtering target position as shown in Fig. 3.

These ROIs also follow the processing way as shown in Fig. 4, which owns K segmentations respectively in order to facilitate local weight distribution under the proposed following structural regression model.

$$G_{n\times p} = [\{X^{positive}_{t+1}\}_{1\times p}, \{X^{negative}_{t+1}\}_{(n-1)\times p}] \quad H_{n\times p} = [\{1\}_{1\times p}, \{-1\}_{(n-1)\times p}] \tag{10}$$

Fig. 6 lists the inputs of vector G in the PLS calculation. Moreover, H input is also labeled {1 or −1} by the same corresponding elements in G as given in (10). The corresponding

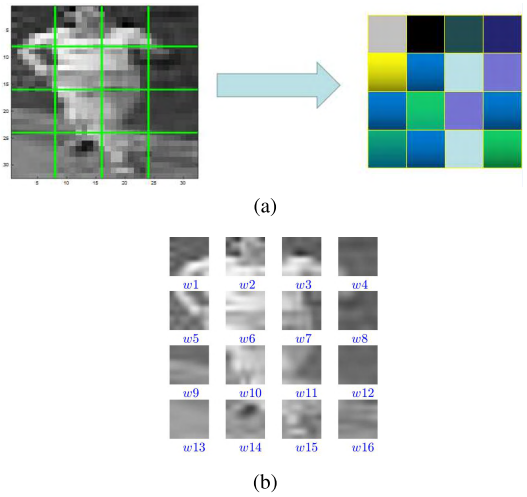


FIGURE 4. K Parts - local patch weights $\{w_i\} \in \{w_1, w_2, w_3, \ldots, w_K\}$. (a) K colorful segments represent different local patch weights. (b) Local weight patch.

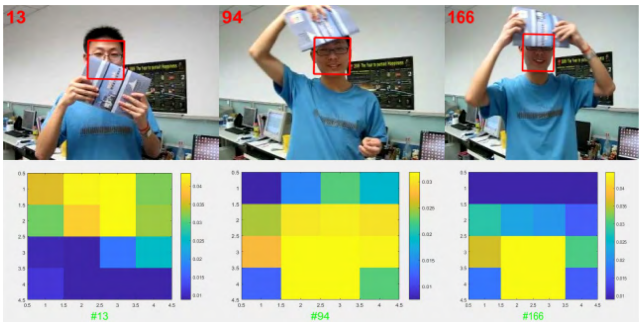


FIGURE 5. Optimal local weight distribution.

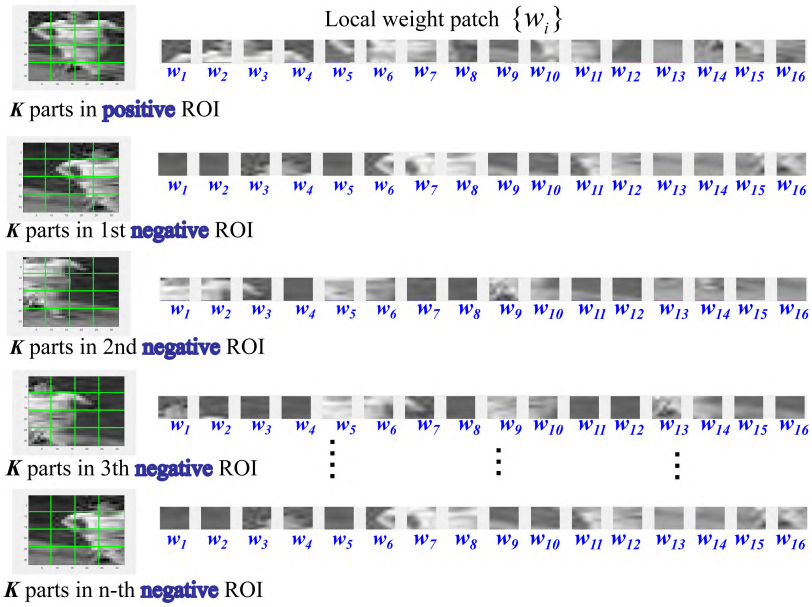


FIGURE 6. PLS inputs vector of *G*:K local weight patches in positive and *n* negative ROIs.

weight for i-th local patch w_i is obtained following the process in (3) and (7) adaptively following the sampling way shown in Fig. 3 and Fig. 6.

Local patch weights $\{w_i\}$ are proposed to facilitate tracking object robustly when appearance varies sharply, such as partial occlusion or deformation. As shown in Fig. 5, small-variation patches take high values in the appearance weight distribution that warmer colors represent the stable local patches. Meanwhile, occlusion or deformation patches are assigned lower weights that present more cool colors.

According to local patch weights $\{w_i\}$ in partial least squares structural regression and the optimal particle states X^*_{t+1} with non-zero coefficients b^*_{t+1} in the inverse sparse representation, the similarity of observation model between the states X_{t+1} and X_t can be estimated easily following the (2).

B. OPTIMAL CANDIDATE SELECTION BY GIST DESCRIPTOR

The optimal state X^{optm} generated by inverse sparse representation with local weight distribution is described by brutal combination of non-zero coefficients b_{t+1} and its corresponding $(t+1)$-th candidate states X^*_{t+1} in the (9). This way may accumulate the tracking deviation due to sparse decomposition in the tracking looping process. In order to represent the generic target more accurately, this paper adopts the anisotropic filter bank to construct Gist descriptor for selection of optimal candidate.

This paper adopts the filter bank with four scales and eight orientations as shown in Fig. 7.

The i-th inverse sparse candidate state X^i_{t+1} in $(t+1)$-th frame is split into a grid on various scales and the output of each cellular grid is calculated using a series of gabor filter

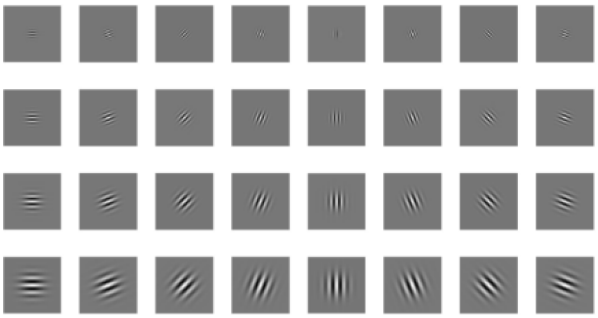


FIGURE 7. Symmetric filter bank with four scales and eight orientations.

bank. Each normalized image of X^i_{t+1} firstly convolves with the symmetric filter bank resulting in thirty-two feature maps of the same size of the input. Then $n \times n$ regions split each feature map that can extract totally $32 \times n^2$ region maps(n^2 regions$\times$32 feature maps). Finally, each region map is averaged and serialized into Gist descriptor that contains $32 \times n^2$ new features $\boldsymbol{\psi}_{\boldsymbol{X}^i_{t+1}}$ in the spatial envelope energy spectrum as shown in Fig. 8. To provide better evaluation between the reference in (t)-th frame and the inverse sparse candidates in $(t+1)$-th frame, this paper adopts the cosine distance to compare the new $32 \times n^2$ features of the reference $\boldsymbol{\psi}_{\boldsymbol{X}^{Ref}_t}$ and the inverse sparse candidates $\boldsymbol{\psi}_{\boldsymbol{X}^i_{t+1}}$ for selecting the most similar candidate X^i_{t+1} as the best one as (11) and (12). Fig. 9 shows the selection for optimal candidate by cosine evaluation.

$$s.t.\ \arg\min \frac{< \boldsymbol{\psi}_{\boldsymbol{X}^{Ref}_t}, \boldsymbol{\psi}_{\boldsymbol{X}^i_{t+1}} >}{||\boldsymbol{\psi}_{\boldsymbol{X}^{Ref}_t}|| \bullet ||\boldsymbol{\psi}_{\boldsymbol{X}^i_{t+1}}||} \tag{11}$$

$$X^{optm}_{t+1} = X^i_{t+1} \tag{12}$$

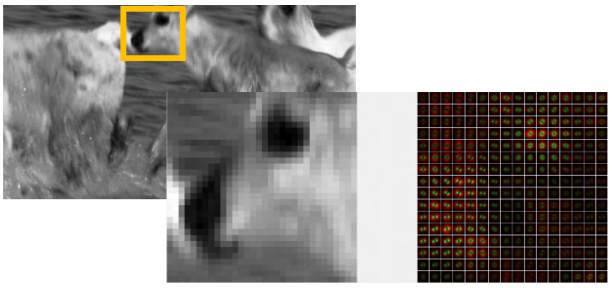


FIGURE 8. Gist descriptor spatial envelope energy spectrum (16 ∗ 16 regions).

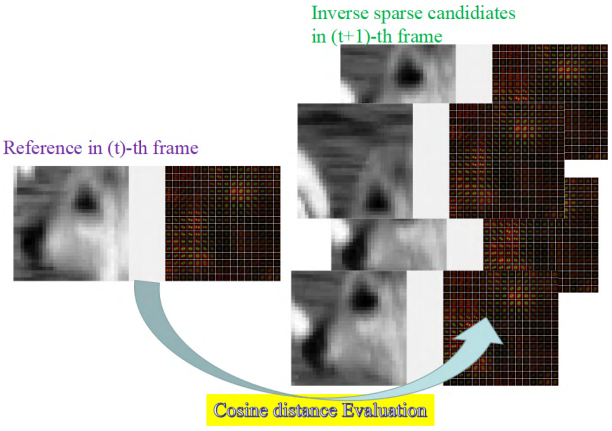


FIGURE 9. Selection for optimal candidate by cosine evaluation (16 ∗ 16 regions).

C. MODEL UPDATE MECHANISM

During the tracking process, it is necessary to update the template t_t adaptively based on the sparse reconstruction error err_{t+1}(13):

$$err_{t+1} = ||t_t - (w_t \bigodot D_{t+1}b_{t+1})||_{l_0} \tag{13}$$

When err_{t+1} is more than a proper threshold τ, the template t_t described in the Section. 3.2 will be iteratively updated by the (14).

$$\begin{aligned} t_{t+1} &= t_t\mu + (1-\mu)(w_t \bigodot D_{t+1}b_{t+1}) \\ s.t.\ b_{t+1} &> 0 \\ s.t.\ err_{t+1} &> \tau \end{aligned} \tag{14}$$

IV. EXPERIMENTAL EVALUATION AND ANALYSIS

The proposed algorithm is implemented using Matlab on a computer with specifications of 2.66-GHz Intel Pentium(R) CPU and 8.0-GB memory. In all the experiments, target areas are selected manually in the first frame image (target ROI) and modeled as above procedures. The L_1 sparse minimization solution is supported by SPAMS package [24] and the regularization constant λ is set to 0.1. Three hundred particles are sampled for providing the candidate sets in each tracking loop.

The proposed algorithm is tested by multiple video clips that contain the nine video sequences [25]. In order to deal

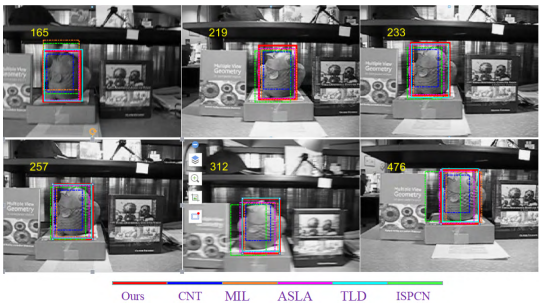


FIGURE 10. Fish tracking results.

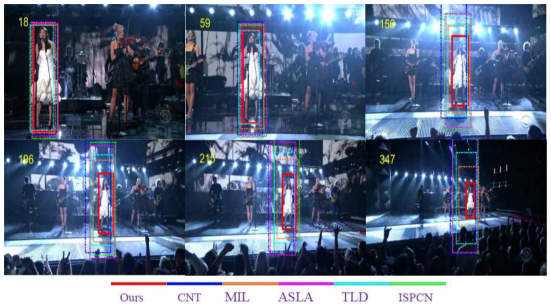


FIGURE 11. Singer1 tracking results.

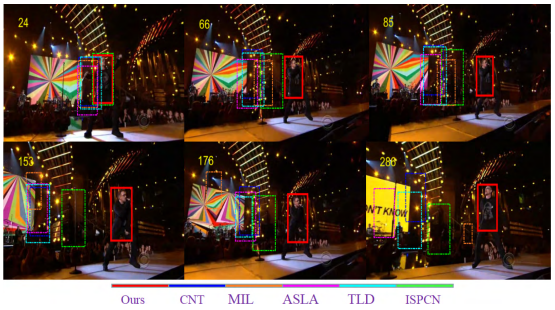


FIGURE 12. Singer2 tracking results.

with the tracking error, the groundtruth of tracking object is labeled manually in each frame. All affine transformations are possessed by 32×32 normalized patch. All pictures are normalized for 640×480 pixels size. The size of local patches is set 8×8 pixels. In order to prove the superiority of the proposed algorithm, this work employs several state-of-the-art challenging algorithms that are related with sparse constraints, including ISPCN(tracker via inverse sparse representation and convolutional networks) [18], CNT(tracker via convolutional networks without training) [17], TLD [26], MIL(multiple instance learning) [27], ASLAS(tracker via structural local sparse appearance) [9].

Fig. 10, 11 and 12 shows that our tracker achieves better results in the illumination variation cases. The good performances are attributed to the Gist descriptor that owns the stable capability of feature description. The proposed algorithm also performs best in spite of the complex background environment such as Fig. 12 that singer walks and sings in the

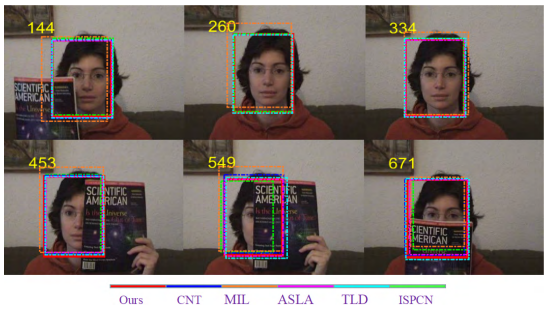


FIGURE 13. Face occlusion1 tracking results.

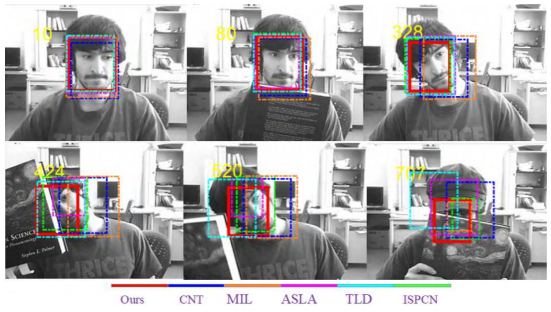


FIGURE 14. Face occlusion2 tracking results.

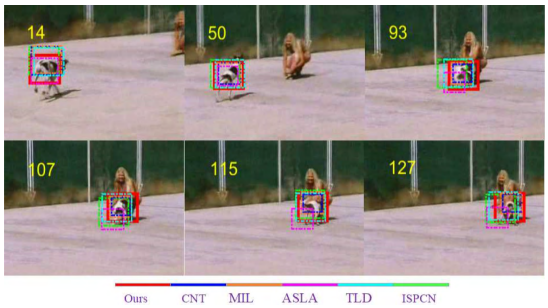


FIGURE 15. Dog tracking results.

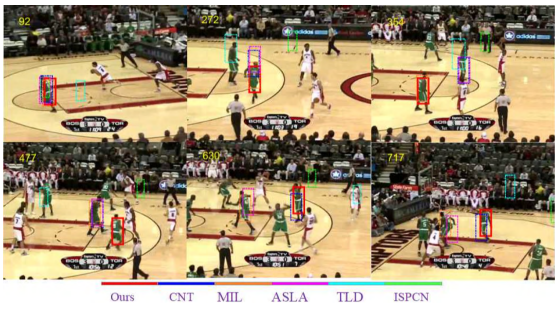


FIGURE 16. Basketball tracking results.

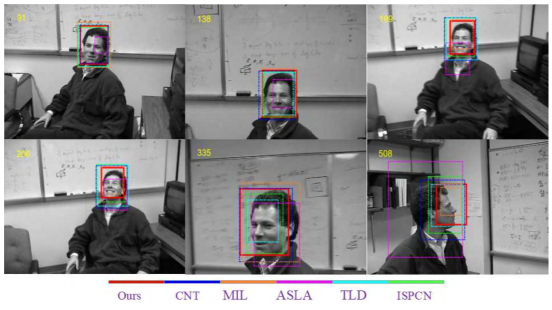


FIGURE 17. Fleetface tracking results.

FIGURE 18. Human2 tracking results.

stage with the severe illumination variation. In addition, there are serious pose variations while the camera focuses on the singer pose cross the whole process. As shown, singer1 video clips are influenced by the stage lights variation in which only our method are able to deal with scale change properly as shown in the Fig. 11. For fish video clips, the fish model undergoes drastic illumination variation with some motion blurring. We can see that ISPCN tracker is not able to track accurately in the Fig. 10. It is easily influenced by the drastic illumination variation easily. The tracking process faces abruptly interruption in almost 275-th frame. However, our tracker always give an important focus on the target, also give the lowest average center error.

Fig. 19.*a*. 19.*b* and 19.*c* give the final corresponding results of fish, singer2 and singer1 respectively. According to the error plot line in Fig. 19, the proposed algorithm also shows the optimal stability.

For occlusion cases(Fig. 13 and Fig. 14), two human faces are all partially occluded by the book. The proposed algorithm obtains good results in this situation. The reasons are as follows :1) the partial least square model in the feature distribution to some extent improves the proposed tracker robustness. 2) the occluded parts are able to be alleviated by only applying the parts without occlusion. In addition, sparse representation based CNT, ISPCN achieve good results by its multi-fusion appearance model. Meanwhile, traditional trackers such as MIL and ALSA are also able to take the unique solutions to occlusion. TLD trackers also give similar performances in this simple case within a certain error range as shown in Fig. 19.*d*. Fig. 14 belongs to heavy occlusion case that has either severe and long-time partial occlusion. Fig. 19.*e* shows the robustness of the proposed algorithm in dealing with rotation and heavy occlusion. Since Gist feature describes the target appearance discriminatively, although particle candidates undergo heavy occlusion, the updated

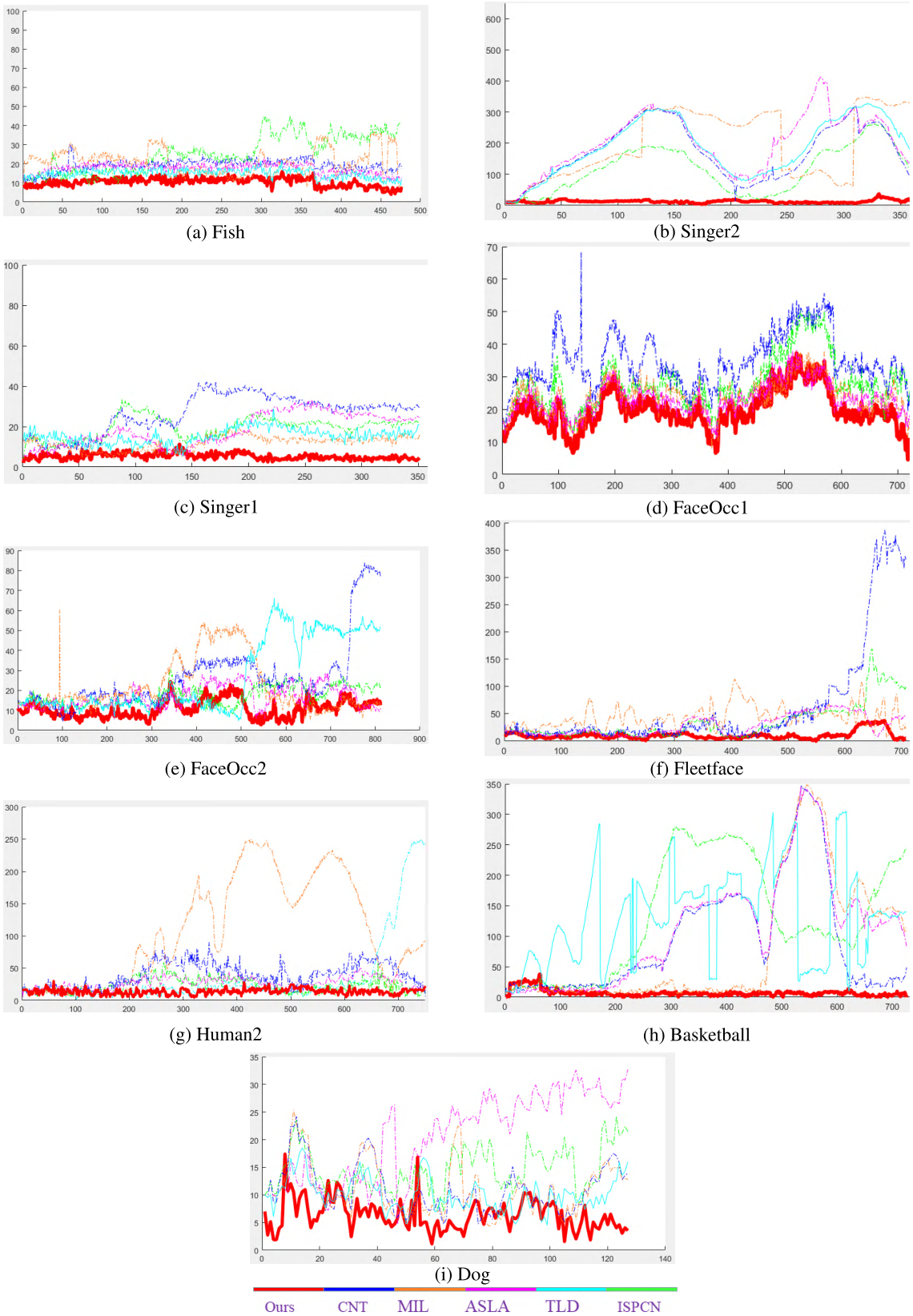


FIGURE 19. Center error representaion. (a) Fish. (b) Singer2. (c) Singer1. (d) FaceOcc1. (e) FaceOcc2. (f) Fleetface. (g) Human2. (h) Basketball. (i) Dog.

TABLE 1. Average center error.

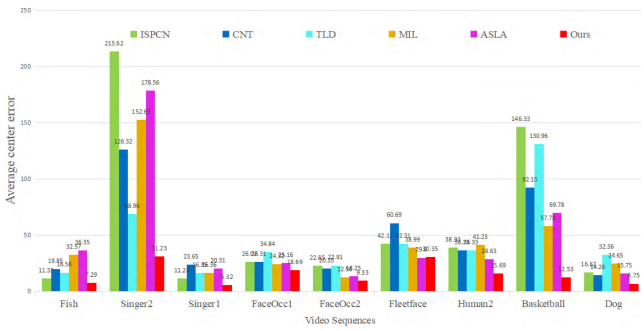
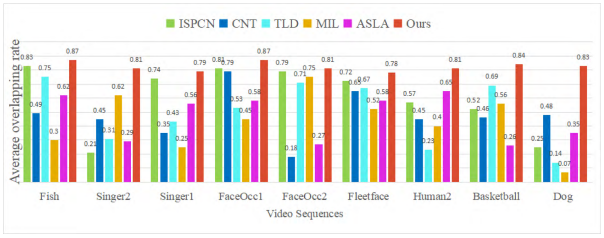


TABLE 2. Average overlapping rate.



templates have no large variation in the global level, which makes inverse sparse coefficients of template higher than that of the bad candidates.

For non-rigid deformation cases(Fig. 16 and Fig. 15), targets undergo abrupt pose variations. It is rather difficult to locate target position when the discriminative appearance always changes in feature distribution. However the proposed algorithm achieves better center error rate than other trackers in Fig. 16. Especially after 180-th frame, TLD, ISPCN, CNT have large error variations but the proposed algorithm still keeps the low error rate. The benefits from the proposed algorithm contain the robustness of model description. Such deformable case often presents low and dense representation errors as shown in Fig. 19.*h*. Fig. 15 shows that the dog target is always running forward the woman. This case has either deformable pose and variable scale in the whole process. The motion variations result in target appearance changing significantly. Because the proposed algorithm adopts online model update strategy, Gist feature excludes the local deformation in the candidate selection. The proposed algorithm achieves a good result in Fig. 19.*i*.

For accidental rotation cases shown in Fig. 17 and Fig. 18, targets face abrupt appearance variations caused by large degree rotation. The proposed algorithm seems effective to deal with such challenging cases. This is the reason that the proposed appearance model is robust and stable against the outliers, Gist feature obtains discriminative identification in the tracking. Fig. 19.*f* and 19.*g* confirm the effectiveness of the proposed algorithm in the long-term tracking. Fig. 18 shows that the proposed algorithm gets lower error comparing with others. Many trackers lose the target when they undergo abrupt pose variations.

In order to evaluate the performance, Tab. 1 and Tab. 2 report the comparisons based on average error rate and average overlapping rate respectively. This way that all the algorithms on the same computer configuration are obtained. The center error rate and the overlapping rate are employed in this section. It is found that the case of low average center error rate and of high overlapping rate will be defined as good performance. As shown in both figures, the proposed algorithm gets favorable performance against others.

V. CONCLUSION

This paper presents the optimal structural regression model based inverse sparse representation for tracking algorithm. Comparing with the other sparse representation based trackers, two key advantages exist in the proposed method. Firstly, considering the potential characteristics of local patches, optimal weights with spatial structural information among local patches are exploited to describe more compact target appearance. To dynamically depict the model appearance, the local weight update scheme in the structural regression process is utilized by the partial least squares regression. This alleviates the drift problem more efficiently and effectively by the proposed method, when the environmental occlusions occur to the tracker. Weights distribution in different patches also facilitates the inverse sparse representation in the tracking process. Explicitly the dictionary construction complexity is avoided for the reason that dictionary is coded by candidate sets of particle samples for the sparse representation. Secondly, incorporating Gist descriptor in frequency domain, optimal set of candidate is more easily selected in global level to jointly capture the target appearance variations. Collections of variant directional wavelet filters (e.g., the gabor feature) are efficiently integrated into the learning process that further improves the discriminative power of optimal candidate selection. Furthermore, the proposed tracker achieves a better performance in the challenging variation environment having illumination with good accuracy. Currently we are working on a new algorithm that merges more discriminitive features which are expected to reduce the calculation cost and improve the robust tracking.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, p. 13, 2006.
[2] F. Porikli and A. Yilmaz, "Object detection and tracking," in *Video Analytics for Business Intelligence*. Berlin, Germany: Springer, 2012.
[3] D. H. Ballard and C. M. Brown, *Computer Vision*, 1st ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1982.
[4] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 125–141, 2008.
[5] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2042–2049.
[6] F. Yang, H. Lu, and M.-H. Yang, "Robust superpixel tracking," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1639–1651, Apr. 2014.
[7] B. Liu, J. Huang, L. Yang, and C. Kulikowsk, "Robust tracking using local sparse appearance model and k-selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1313–1320.
[8] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1838–1845.

[9] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1822–1829.
[10] L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1910–1917.
[11] X. Mei and H. Ling, "Robust visual tracking using ℓ_1 minimization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1436–1443.
[12] C. Sun, D. Wang, and H. Lu, "Occlusion-aware fragment-based tracking with spatial-temporal consistency," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3814–3825, Aug. 2016.
[13] Q. Wang, F. Chen, W. Xu, and M.-H. Yang, "Object tracking via partial least squares analysis," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4454–4465, Oct. 2012.
[14] T. Bai, Y. F. Li, and Y. Tang, "Structured sparse representation appearance model for robust visual tracking," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 4399–4404.
[15] D. Wang, H. Lu, Z. Xiao, and M.-H. Yang, "Inverse sparse tracker with a locally weighted distance metric," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2646–2657, Sep. 2015.
[16] L. Wang, W. Ouyang, X. Wang, and H. Lu, "Visual tracking with fully convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 3119–3127.
[17] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, "Robust visual tracking via convolutional networks without training," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1779–1792, Apr. 2016.
[18] H. Wang and H. Ge, "Object tracking via inverse sparse representation and convolutional networks," *Optik-Int. J. Light Electron Opt.*, vol. 138, pp. 68–79, Jun. 2017.
[19] A. Heyden and M. Pollefeys, "Multiple view geometry," *Emerg. Topics Comput. Vis.*, vol. 2, nos. 9–10, pp. 45–107, 2005.
[20] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Proc. Int. Conf. Subspace, Latent Struct. Feature Selection*, 2005, pp. 34–51.
[21] Wikipedia Contributors. (2018). *Partial Least Squares Regression—The Free Encyclopedia*. Accessed: Sep. 13, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Partial_least_squares_regression&oldid=863508939
[22] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
[23] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, no. 10, pp. 847–856, 1980.
[24] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Found. Trends Comput. Graph. Vis.*, vol. 8, nos. 2–3, pp. 85–283, 2014.
[25] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 2411–2418.
[26] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.
[27] B. Babenko, M.-H. Yang, and S. J. Belongie, "Visual tracking with online multiple instance learning," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Miami, FL, USA, Jun. 2009, pp. 983–990.

XIAOWEI AN received the B.S. degree in electronics engineering from the Shandong University of Science and Technology, Qingdao, China, in 2008, and the M.S. degree in electronics engineering in 2011. He is currently pursuing the Ph.D. degree with the Shandong University of Science and Technology. His research interests include signal processing and vision pattern analysis.

NONGLIANG SUN received the B.S. degree in physics from Nankai University, China, in 1986, and the M.S. and Ph.D. degrees in control theory and control engineering from the Shandong University of Science and Technology (SDUST), China, in 2002 and 2007, respectively. He was a Visiting Scholar in Germany, South Korea, and USA. He is currently the Vice Dean of the College of Electronics, Communication and Physics, SDUST. He has led one 863 Sub-project and participated in several Natural Science Foundations of China and Shandong province. He has published more than 70 papers and several monographs. His research interests mainly focus on computer vision, supersonic detection, automatic control, and virtual reality.

MAOYONG CAO received the bachelor's degree from Nankai University, China, in 1986, the M.S. degree from the Shandong University of Science and Technology, China, in 1993, and the Ph.D. degree from Tianjin University, China, in 2002. He is currently a Professor with the College of Electrical Engineering and Automation, Shandong University of Science and Technology, and a Primary Committee Member of the State Key Laboratory for Mining Disaster Prevention and Control, co-founded by Shandong Province and the Ministry of Science and Technology. He has published over 100 research papers in international journals and conferences, has authored three books, and has been granted 16 patents. His major research interests include ultrasonic testing, image processing, and control engineering.